N95- 33286

DATA COLLECTION IN IMS : IT'S NOT AS EASY AS IT LOOKS

Dennis M. Davis[1], Waleed Maswadeh[2], Donald B. Shoff[1], Charles S. Harden[1], A. Peter Snyder[1]

[1]U.S. Army Edgewood Research Development and Engineering Center, Lightweight Chemical Detector Team, Aberdeen Proving Ground, MD 21010-5423, [2]GEO-CENTERS, Inc., P.O. Box 68, Gunpowder Branch, Aberdeen Proving Ground, MD 21010-5423

ABSTRACT

Data collection in Ion Mobility Spectrometry (IMS) is not as easy an endeavor as it appears. Despite the advent of high speed personal computers and fast analog-to-digital converters (ADC's), care must be taken to ensure that reliable data is obtained in a timely fashion. This is especially true in hyphenated techniques, e.g. GC-IMS, where the amount of data increases dramatically when gas chromatography-ion mobility spectrometry (GC-IMS) data is being collected. Using the Graseby GC-IMS, with a gate repetition rate of 33 Hz, it is theoretically possible to collect 33 spectra per second. This collection rate is not realistically obtained due to a number of factors. Among these factors are inaccuracy of the timing signal from the IMS, the necessity to store the data, disk input/output limitations, disk operating system limitations, and program overhead. Taking these factors into account, we have achieved a data collection rate of 20 spectra per second. This paper will describe these problems, demonstrate the practical effects these problems present, and present methods for minimizing these effects.

INTRODUCTION

One of the most important aspects of ion mobility spectrometry (IMS) studies, and too often the least understood, is data collection and data storage. Detection, identification and determination of the concentration of the chemical species present are all affected by the data collection. Yet, as important as data collection is, many researchers do not take the time to collect and store IMS data properly, either because they use routines written by others, or they are not fully aware of the limitations under which they must operate. The limitations may be classified into one of three categories: hardware related, software related, or data related. The limitations which most affect the data collection and data storage are as follows:

Hardware related:
(1) inaccuracy of the timing signal from the IMS,
(2) disk input/output limitations,
(3) computer limitations,

software related:
(4) disk operating system limitations,
(5) program overhead, and

data related:
(6) the necessity to store the data,
(7) the nature of the data itself.

EXPERIMENTAL

Hardware.

For this work, we used an improved Environmental Vapor Monitor, EVM, (Graseby Ionics, Ltd. Watford, Herts, UK). The EVM, a hand-held GC-IMS, is comprised of capillary gas chromatograph, GC, integrated with a hand-held ion mobility spectrometer, IMS[1]. The IMS operates with an internal sample gate repetition rate of 33 Hz. The gating pulse has an amplitude of 5 volts and is 180 microseconds in duration. This gating pulse provides the trigger for the data collection. The improvements include the introduction of temperature and pressure sensors inside the IMS cell and the construction of a disposable GC module. The disposable GC module offers some important features: 1) an easily replaceable GC module which facilitates changes in column types and lengths,as dictated by the compounds being studied, 2) easy replacement of the GC column when the column is spent, and 3) a GC column that can be easily heated from room temperature to 130°C in two minutes. The improvements in the GC-IMS system design ensure better separation characteristics, improved detection and identification of chemical compounds, increased ease of maintenance of the system, and a more robust hand-held detector. Typical experimental conditions used for the hand-held GC-IMS are shown in Table I. Sample introduction to the GC column was accomplished by using an Automated Vapor Sampling unit, AVS[2]. The sample pulse is user controlled with a range of 0.2 seconds to 2 seconds duration.

TABLE I
EXPERIMENTAL OPERATING CONDITIONS OF THE GC-IMS

Disposable GC Module:		
GC Column:	**Liquid Phase:**	**DB-1 (0.25 micrometer I.D.)**
	Temperature (°C):	**45°C/min programmable**
	Carrier Gas:	**Clean dry air**
	Flow Rate:	**2.1 ml/min**
	Length:	**1 m**
Sample Injection:	**User-software controlled typically 0.2 sec duration**	
Ion Mobility Spectrometer:		
Ionization Source:	**^{63}Ni**	
Gating Pulse Repetition Rate:	**33 Hz**	
Cell Temperature:	**30 °C**	
Cell Pressure:	**640 torr**	
Drift Gas:	**Clean dry air**	
Drift Gas Flow:	**400 ml/min**	

Data was collected on a Dell 486D personal computer operating at a processor speed of 33MHz, with 8 KB of internal cache memory, 32 MB of system memory (RAM), a 230 MB IDE hard disk drive with 16 ms average seek time, and VGA monitor. The data was collected using an AT-MIO-16X multifunction I/O board (National Instruments Corporation, Austin, TX). The AT-MIO-16X has a 10 microsecond, 16 bit, sampling ADC. Typical data collection parameters are shown in Table II.

TABLE II.
TYPICAL DATA COLLECTION PARAMETERS USED WITH THE GC-IMS

Ion Mobility Spectrometer Parameters:	
Spectral Mode:	Positive Ion
Points Per Spectrum:	640
Sampling Frequency:	33 Khz
Delay to Start of Sampling:	0 microseconds
GC Parameters:	
GC Delay:	0 seconds
GC Heating Time:	30 seconds
AVS Temperature:	Room
AVS Pulse Duration:	100 ms
Data Acquisition Board Parameters:	
PC Slot Number:	1
Digital I/O Port Address:	A
IMS Analog Input Channel:	0
IMS Mode Detect Channel:	0
GC Digital Trigger Channel:	1
Cell Temperature Input Channel:	4
AVS Temperature Input Channel:	5
Column Temperature Input Channel:	6
Cell Pressure Input Channel:	3

Software.

The data collection algorithms were written using Labwindows Software Version 2.2 (National Instruments, Austin,TX) in both C and QuickBasic programming languages. The programs were compiled using Microsoft C Version 5.1 and Microsoft QuickBasic as appropriate. The compiled versions of the software were then run under Microsoft DOS Version 6.0. Labwindows versions of the executable code were all created using the Labwindows Run Time System.

DISCUSSION

Hardware Related Limitations.

Of all the limitations, the hardware related limitations are the ones over which the researcher has the least direct control. The most severe limitations are caused by inaccuracy of the timing of the gating pulse, disk input/output limitations, and computer limitations. Each of these may be overcome to some extent, but each extracts a price.

Inaccuracies in the timing signal from the IMS. The gating pulse on the IMS is used as the trigger to start the data collection. The accuracy of the timing of the gating pulse is directly related to the accuracy of the clock used to time the pulse. While the gating pulse repetition rate of the GC-IMS is 33 Hz, there are inconsistencies in the gating pulse rate as shown in Figure 1. These inconsistencies, or inaccuracies, affect the data collection. The data collection is not affected by the gating pulse repetition rate being slower than the 33 Hz as much as it is when the rate is faster than 33 Hz. Since 640 data points are collected at a frequency of 33 kHz, the total time necessary to collect one spectrum is 21.12 milliseconds. At a gating pulse repetition rate of 33 Hz, this leaves 8.88 milliseconds between the end of data collection and the start of the next spectrum to write the data to the disk. Slowing of the data gating pulse repetition rate provides more time to complete the transfer of data to the disk. Thus, the effect of slowing the data rate by a slight amount is negligible. However, when the data rate is increased slightly, the amount of time available to transfer the data to the disk is decreased. Thus, the chance of getting a trigger pulse to start data collection during the data transfer to disk is increased. Since the program has not issued the commands to initiate the data collection, the spectrum is lost and the program waits for the next trigger pulse. Thus, it is possible to envision a decrease in the sampling rate of 33 Hz to 15 Hz without taking other limitations into account.

A sequence of ion mobility scans is shown in Figure 2 and demonstrates another problem which may be encountered. It is noted that although the repetition rate of the gating pulse and data is approximately 33 Hz, there are periods in which no trigger signal, and thus, no data is available. The GC-IMS produces eight trigger pulses and spectral data sets, skips four pulses and data sets, then repeats the pattern. The source of this periodicity is internal to the GC-IMS, and therefore beyond control of the experimenter. This problem may be unique to those IMS devices which are designed to provide averaged spectra. In actuality, with the GC-IMS, there are at most 24 spectra available for collection per second. The effect that the non-uniformity of the gating pulse has on the data collection is shown in Figure 3, which is a contour plot of the first 11 spectra collected in a GC-IMS run, with each spectrum in Figure 3 being represented by a bar. The GC retention time reflects the inaccuracy in the timing of the gating pulse. There should be consistent spcaing, retention time difference, between consecutive bars in Figure 3. This problem may be overcome by issuing the trigger pulse from the computer rather than the IMS. This generally entails a reworking of the IMS electronics. If one uses more than one IMS, the time and energy spent on this solution becomes burdensome.

Disk input/output limitations. The second of the hardware limitations, and the least understood, are the disk input/output limitations. These limitations include disk access time, disk cluster size, and disk access procedures. Typically, disk access times range from 13 to 21 milliseconds, with the time being a function of disk size. Some of the more common drives and their disk access times are given in Table III. The data in Table III was compiled from a variety of sources including manufacturer data sheets, computer system documentation, and computer shopper magazines. A complete description of drive types may be found elsewhere[3]. To improve the speed of the disk drive, one must usually increase the size of the drive, resulting in increased computer costs. However, when speed is of the essence, it is a price well spent.

The next disk input/output limitation, disk cluster size, is more insidious than the other limitations, because it is generally hidden from the programmer. The disk cluster size is a function of the disk size, as shown in Table IV. While the disk cluster size does not directly affect the speed of data collection, its effects are nevertheless present. For example, when collecting individual IMS spectra, which have a typical size of 1300 bytes, collecting a large number of individual spectra on a 200 MB drive results in a waste of disk space of almost 75 percent. Thus, when each subsequent spectrum is saved to disk, there are fewer clusters in which to save data, the clusters are generally scattered across the surface of the disk, and the disk access time slows as free clusters are located. This problem is easily overcome by collecting all spectra to a single spectral file, thus ensuring that at most 1 disk cluster is lost

TABLE III.
DISK TYPES AND ACCESS TIMES

MANUFACTURER	DRIVE TYPE	DRIVE SIZE	DISK ACCESS TIME
Conner	SCSI	212 MB	12 ms
Conner	SCSI	170 MB	17 ms
Conner	IDE	42 MB	25 ms
Conner	IDE	212 MB	12 ms
Seagate	IDE	43 MB	28 ms
Seagate	IDE	245 MB	12 ms
Maxtor	IDE	213 MB	12 ms
Maxtor	SCSI	213 MB	15 ms

for a set of spectra. The cost for this solution is simply that individual spectra are more difficult to access. This cost is more than offset by the increased disk storage capacity and time saved in saving spectra.

TABLE IV.
DISK CLUSTER SIZES

Disk Size		DOS Default Cluster Size
Floppy Disks:		
360	KB	1024 Bytes
1.2	MB	512 Bytes
1.44	MB	512 Bytes
Hard Disks:		
0-16	MB	4096 Bytes
16-128	MB	2048 Bytes
128-256	MB	4096 Bytes
256-512	MB	8192 Bytes
512-1024	MB	16,384 Bytes
1024-2048	MB	32,768 Bytes

Computer limitations. The easiest of the hardware limitations to overcome are the computer limitations. The limitations in this category are related to the speed at which instructions may be carried out by the computer: processor speed limitations. Upgrading computer facilities is the only solution available. Some of the possible components which may be included in this upgrade are:

1) install a direct memory access (DMA) board,
2) install a newer I/O board,
3) replace the current computer with one which has a faster processor.

Generally, replacing the computer with one which has a faster processor will have the greatest impact on the speed of accessing and saving data. Installing a DMA board will improve the speed of the data transfer from memory to disk, but has little impact on how fast the other instructions are performed. A newer I/O board will allow faster data collection from the IMS, but it won't improve the speed of the data transfer to disk, or perform other instructions faster. Improving the processor speed will improve the speed at which each instruction is performed, which minimizes the total time necessary to complete the task of data transfer and manipulation.

Software related limitations.

There are two basic types of software related limitations. The first are those related to the disk operating system, DOS, and the second are those related to program overhead. These limitations are generally outside the control of the researcher.

Disk operating system limitations. The most important function of DOS is the control of access to the disks in the computer. Disk access and file allocation of disk space is performed as requested by the program, with space being allocated one cluster at a time. The allocation algorithm used by DOS is called the Next Available Cluster algorithm[3]. Each time the program sends a command to write to the disk, the algorithm starts at the cluster where the last write occurred, and then searches for the next free cluster to begin writing the data. After that cluster is written, the next free cluster is located, and the write continues until the data is written. Because the next available cluster is used, the data file may be spread over a large amount of the disk, depending upon where the next available cluster is located, and the file becomes fragmented. In addition to writing the data, the computer's operating system must place the pointers of which clusters belong to which files in the File allocation Table, FAT, and place a directory listing for the file in the proper directory listing. If duplicate filenames are used within the same directory, the current directory listing and FAT pointers must be removed from the FAT and the new listings and pointers entered. This results in increased time required to write successive files, as shown in Table V, where the number of files written in each second is seen to decrease as the number of files increases.

The solution to this limitation is again to write the data to a single data file for each set of spectra collected. Thus, the next available cluster will often be the current cluster (until it is filled with data). The directory listing must only be made once, and only the FAT listing must be updated.

Program overhead limitations. Program overhead limitations may be imposed upon the researcher and programmer by DOS, as seen above, or by the programmer himself. The limitations that the programmer places on the data collection are related to such mundane tasks as: reading the system clock to determine how much time has expired since the last spectrum was collected, keeping track of how many spectra have been collected to date, monitoring for trigger pulses, and maintaining a check on the status of data transfers to the disk. Each of these steps may be eliminated, but the price which must be paid is an increased level of uncertainty as to the nature of the data. For example, failure to read the system clock will result in not knowing when a particular sample was collected. Thus, its place in the series of spectra which have been collected is unknown. Not monitoring for trigger pulses is a particularly unpleasant idea when taking GC-IMS data, because the GC retention time is unknown, and the information gained from performing the GC separation is lost. Monitoring the status of the data transfer to disk is the most dangerous of the program overhead steps to eliminate, because it is not just information you may lose, you run the risk of losing the data itself.

The greatest burden on program overhead is the desire of the researcher to view the data as it is being collected. This necessitates the programmer putting in graphics routines. Graphics routines generally require additional programming steps related to scaling the data to fit the graphics windows and

TABLE V.
PARTIAL DIRECTORY LISTING FOR SUCCESSIVELY CREATED SPECTRA

TIME0000.ACQ	1334	10-03-94	1:41p
TIME0001.ACQ	1334	10-03-94	1:41p
TIME0002.ACQ	1334	10-03-94	1:41p
TIME0003.ACQ	1334	10-03-94	1:41p
TIME0004.ACQ	1334	10-03-94	1:41p
TIME0005.ACQ	1334	10-03-94	1:41p
TIME0006.ACQ	1334	10-03-94	1:41p
TIME0007.ACQ	1334	10-03-94	1:41p
TIME0008.ACQ	1334	10-03-94	1:41p
TIME0009.ACQ	1334	10-03-94	1:41p
TIME0010.ACQ	1334	10-03-94	1:41p
TIME0011.ACQ	1334	10-03-94	1:41p
TIME0012.ACQ	1334	10-03-94	1:41p
TIME0013.ACQ	1334	10-03-94	1:41p
TIME0014.ACQ	1334	10-03-94	1:41p
TIME0015.ACQ	1334	10-03-94	1:41p
TIME0016.ACQ	1334	10-03-94	1:41p
TIME0017.ACQ	1334	10-03-94	1:41p
TIME0018.ACQ	1334	10-03-94	1:41p
TIME0019.ACQ	1334	10-03-94	1:41p
TIME0020.ACQ	1334	10-03-94	1:41p
TIME0021.ACQ	1334	10-03-94	1:41p
TIME0022.ACQ	1334	10-03-94	1:41p
TIME0023.ACQ	1334	10-03-94	1:42p
TIME0252.ACQ	1334	10-03-94	1:48p
TIME0253.ACQ	1334	10-03-94	1:49p
TIME0254.ACQ	1334	10-03-94	1:49p
TIME0255.ACQ	1334	10-03-94	1:49p
TIME0256.ACQ	1334	10-03-94	1:49p
TIME0257.ACQ	1334	10-03-94	1:49p
TIME0258.ACQ	1334	10-03-94	1:49p
TIME0259.ACQ	1334	10-03-94	1:49p
TIME0260.ACQ	1334	10-03-94	1:49p
TIME0261.ACQ	1334	10-03-94	1:49p
TIME0262.ACQ	1334	10-03-94	1:49p
TIME0263.ACQ	1334	10-03-94	1:49p
TIME0264.ACQ	1334	10-03-94	1:49p
TIME0265.ACQ	1334	10-03-94	1:49p
TIME0266.ACQ	1334	10-03-94	1:50p
TIME0357.ACQ	1334	10-03-94	1:57p
TIME0358.ACQ	1334	10-03-94	1:58p
TIME0359.ACQ	1334	10-03-94	1:58p
TIME0360.ACQ	1334	10-03-94	1:58p
TIME0361.ACQ	1334	10-03-94	1:58p
TIME0362.ACQ	1334	10-03-94	1:58p
TIME0363.ACQ	1334	10-03-94	1:58p
TIME0364.ACQ	1334	10-03-94	1:58p
TIME0365.ACQ	1334	10-03-94	1:58p
TIME0366.ACQ	1334	10-03-94	1:58p
TIME0367.ACQ	1334	10-03-94	1:59p
TIME0598.ACQ	1334	10-03-94	2:34p
TIME0599.ACQ	1334	10-03-94	2:35p
TIME0600.ACQ	1334	10-03-94	2:35p
TIME0601.ACQ	1334	10-03-94	2:35p
TIME0602.ACQ	1334	10-03-94	2:35p
TIME0603.ACQ	1334	10-03-94	2:35p
TIME0604.ACQ	1334	10-03-94	2:36p
TIME0605.ACQ	1334	10-03-94	2:36p
TIME0606.ACQ	1334	10-03-94	2:36p
TIME0607.ACQ	1334	10-03-94	2:36p
TIME0608.ACQ	1334	10-03-94	2:36p
TIME0609.ACQ	1334	10-03-94	2:37p
TIME0610.ACQ	1334	10-03-94	2:37p
TIME0611.ACQ	1334	10-03-94	2:37p
TIME0612.ACQ	1334	10-03-94	2:37p
TIME0613.ACQ	1334	10-03-94	2:37p
TIME0614.ACQ	1334	10-03-94	2:38p

refreshing the screen when the next data set is collected. The time necessary to perform these steps increases in direct proportion to the number of points being displayed in the graphics routine. While it is often desirable to monitor the data as it being collected, it is a step which must be eliminated when maximizing the rate at which spectra are collected.

It becomes apparent that program overhead has many items which may be eliminated easily, but the price for eliminating these items is steep indeed. It is possible to minimize the number of times that you perform the monitoring procedures, but they cannot be eliminated. They are simply the price that must be paid to collect data. A realistic determination of what is important to monitor must be made before finalizing the data collection routines.

Data related limitations.

The data related limitations are both the easiest and most difficult limitation to deal with. The limitations are the necessity to store the data, and the nature of the data itself. The researcher has total control over each of these limitations. He may decide that a particular set of data, or portion of a GC run, contains no useful information, and thus doesn't need to be saved. He must accept the fact, however, that he may be wrong and may lose some important data. This loss of data is dependent upon the researcher's knowledge and experience. It is possible that a researcher only wishes to perform a cursory scan of the data; to get an idea of what information is available from the sample. It is quite possible, even probable, that there is no necessity to store the data; the result being that speed of the data collection is increased.

The researcher also has control over the nature of the data. This does not mean that one has control over the IMS spectrum, just what information is germane. There may be only a window of data that is of importance to collect and save. In the GC-IMS, it may be a window of GC retention times, a window of IMS drift time, or both. This is dependent upon the nature of the information that the researcher is attempting to obtain. There is other data which may be important to the researcher as well. A determination of what additional information is important, whether or not to collect the information, and how often to collect the information must be made. Information of this type includes the date and time the data is collected, the GC retention time, the temperature of the locale where the data is collected, the temperature of the GC inlet, the temperature of the GC column, and the temperature and pressure of the IMS cell.

RESULTS

After taking all of these limitations into account, we have developed a software package which is capable of collecting up to 20 spectra per second, or 83 percent of the spectra which are produced by the GC-IMS. Along with the spectra, we monitor and collect the information contained in Table VI.

To achieve the 20 spectra per second rate, it was necessary do eliminate the graphics procedures. Using graphics displays to view data while it is being collected, reduces the rate at which data can be collected to 7 spectra per second. The elimination of the graphics does not prohibit viewing of the data; it only delays the viewing until after all the data has been collected. A typical spectrum obtained using this new data collection routine is shown in Figure 4. The spectra in Figure 4 are plotted in contour format to ease in the visualization of the data.

CONCLUSIONS

At first glance, collecting data from a GC-IMS appears to be a simple task; turn on the computer and the GC-IMS, then take data. It is possible to collect the GC-IMS data at 33 spectra per second, or at least the 24 spectra per second that the GC-IMS produces, but important information is lost and other information is hidden. Information that is hidden is the elapsed time since the AVS sampled the environment. This information is hidden only if there was one sampling pulse at the beginning of the data collection, otherwise it is lost. Information that is lost are cell temperature, cell pressure, AVS temperature, and GC temperature. Thus, the simplistic approach to data collection must be abandoned in the face of reality. There are limitations to be addressed at every turn: limitations related to hardware, software, and data. Choices must be made as to what information is to be retained, and is superfluous to the data. The more aware a researcher is of the limitations, the better job he can do on collecting the data.

TABLE VI
PARAMETERS COLLECTED WITH A GC-IMS SYSTEM, AND THE FREQUENCY WITH WHICH THEY ARE COLLECTED

Parameters collected only at the beginning of the GC-IMS data collection.	
IMS Mode	- Positive or negative ions
Comment	- Information about the sample, column type, etc.
Filename	- Name under which the data is stored
Length of filename	- used to control how data is written to the datafile
Length of comment	- used to control how data is written to the datafile
RATE	- The rate at which data points in a spectrum will be collected
NOSAMP	- Number of points in each spectrum collected
Parameters collected or monitored with each spectrum	
Count	- How many spectra have been collected
TEMPAVS	- Temperature of the AVS Inlet
TEMPCOL	- Temperature of the GC Column
TEMPCELL	- Temperature of the IMS Cell
PRESSCELL	- Pressure inside the IMS Cell
DURATION	- GC Retention Time
PULSE	- Is the AVS actively Sampling?
DISKFULL	- Is the Disk full ?

REFERENCES

1. Snyder, A.P., Harden, C.S., Brittain, A.H., Kim, M.-G., Arnold, N.S., and Meuzelaar, H.L.C. Portable Hand-Held Gas Chromatography/Ion Mobility Spectrometry Device. *Anal. Chem,* **65**, 299-306 (1993).

2. Dworzanski, J.P., Kim, M.-G., Snyder, A.P., Arnold, N.S., and Meuzelaar, H.L.C. Performance advances in ion mobility spectrometry through combination with high speed vapor sampling, preconcentration and sepaeation techniques. *Anal. Chim. Acta*, **293**, 219-235 (1994).

3. Mueller, S. and Elliot, A.C. Que's Guide to Data Recovery, Que, Carmel, IN, 1991, pp 76-84.

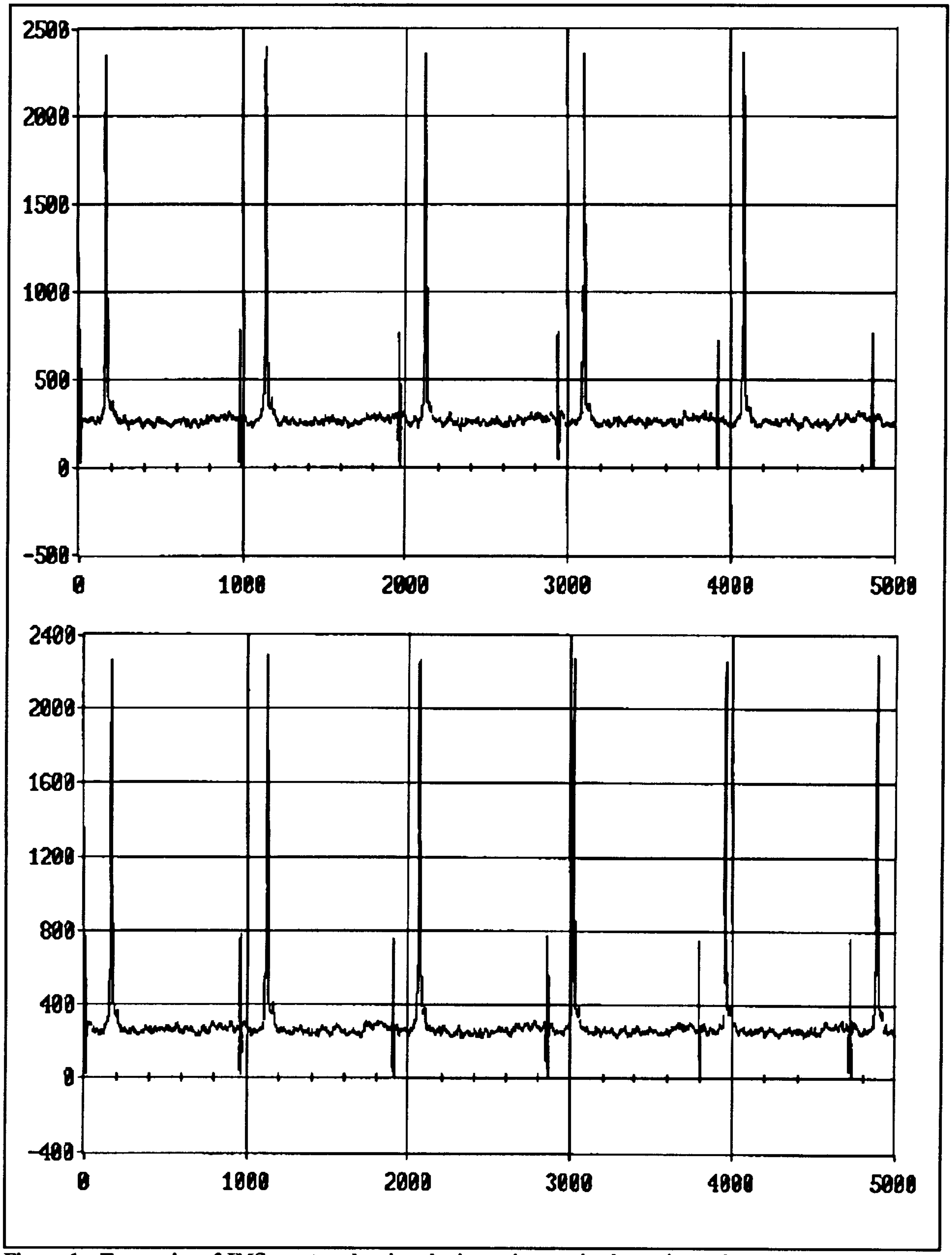

Figure 1. Two series of IMS spectra showing the inconsistency in the gating pulse rate.

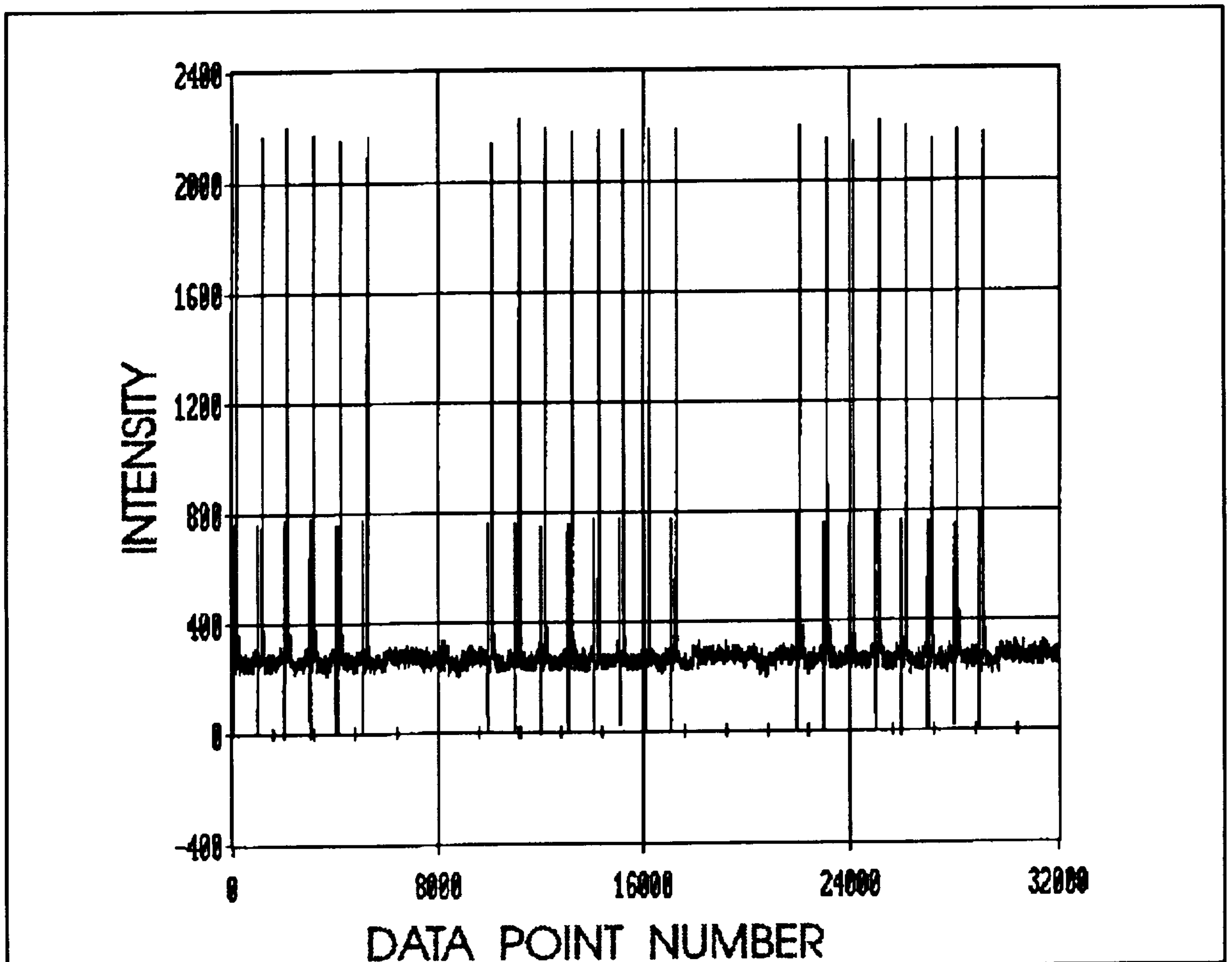

Figure 2. IMS spectra collected over a one second interval. Thirty thousand data points were collected at a data rate of 30,000 Hz.

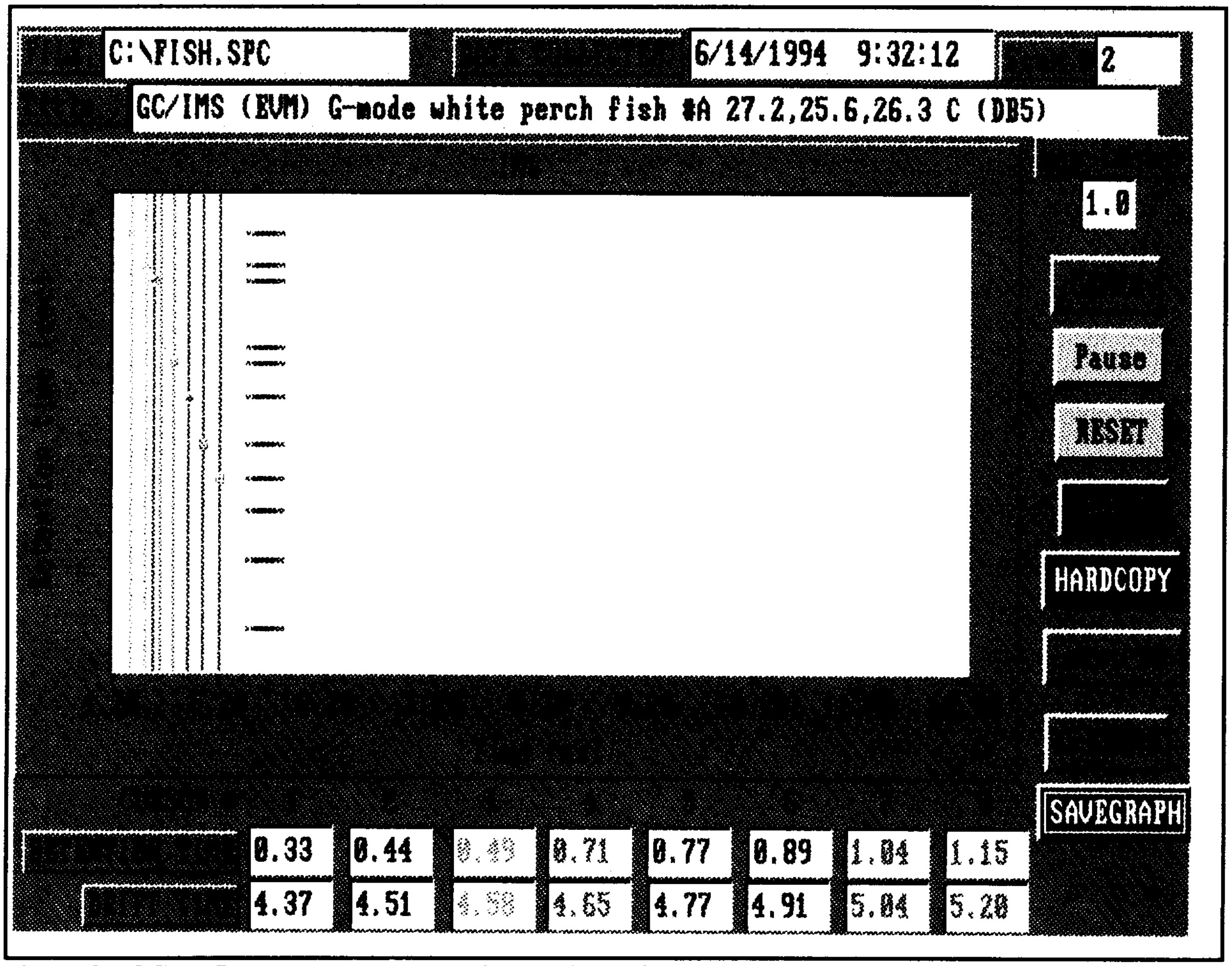

Figure 3. GC-IMS contour spectra showing the inconsistency of the gating pulse, as indicated by the GC retention times of each spectrum.

C-4.

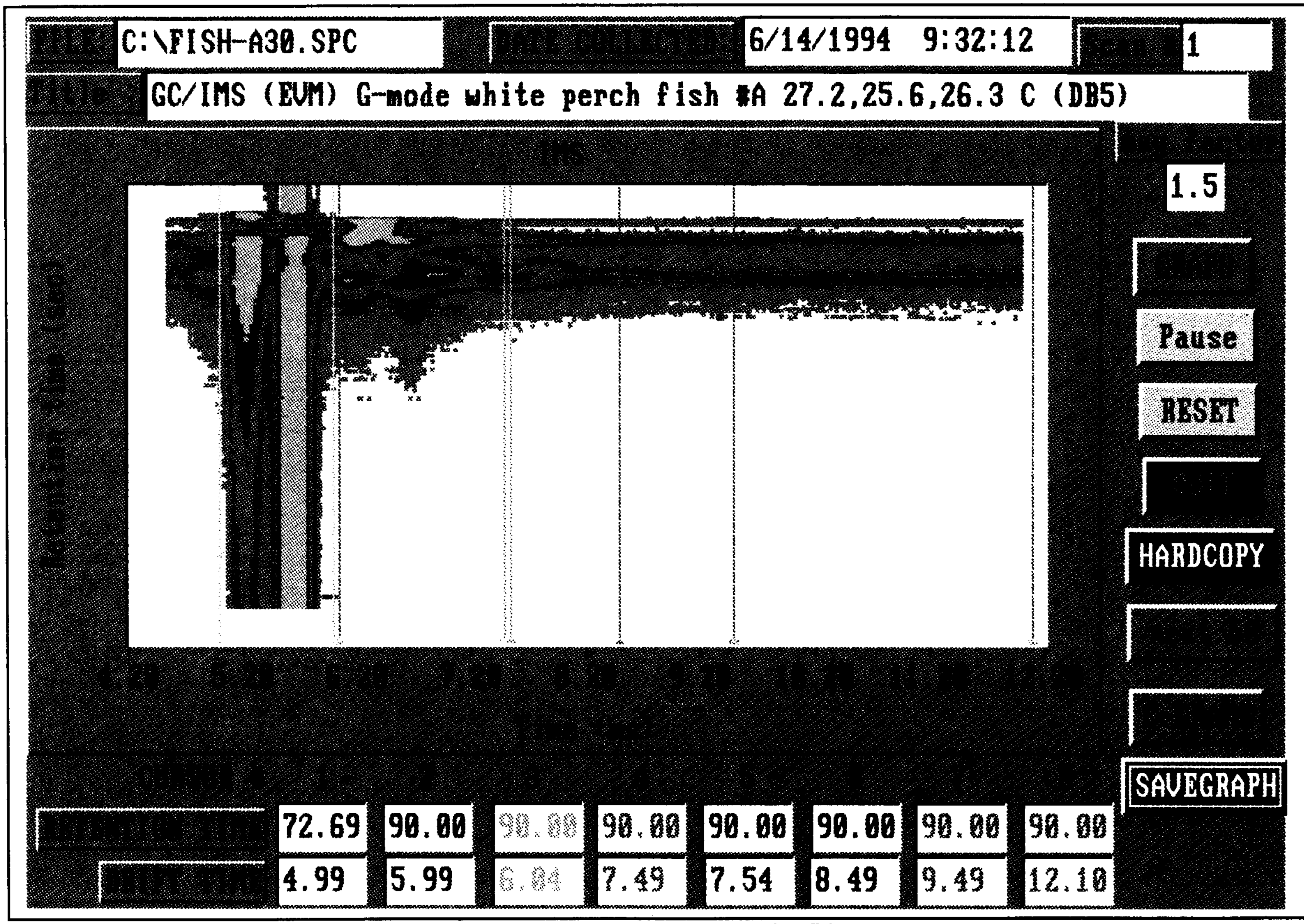

Figure 4. Contour plot of the GC-IMS spectra collected from a sample of decomposing fish.